



हिंदी शब्दतंत्र की संरचना

प्रभाकर पाण्डेय

prabhakargopalpuriya@gmail.com

आईआईटी मुंबई की सेवा के बाद सी-डैक पुणे में राजभाषा अधिकारी

लक्ष्मी कश्यप

yupu@cse.iitb.ac.in

आईआईटी मुंबई में कार्यरत

पुष्पक भट्टाचार्य

pushpakbh@gmail.com

भाषा-प्रौद्योगिकी के विशेषज्ञ एवं आईआईटी मुंबई में प्रोफेसर

शब्दतंत्र (Wordnet) किसी भी भाषा का एक बहुत बड़ा शब्द संचय (Database) होता है, जिसमें संज्ञा, विशेषण, क्रिया और क्रिया-विशेषण संज्ञानात्मक समानार्थक शब्दों या पर्यायों को समूह में एकत्रित किया गया होता है और जिनमें से प्रत्येक शब्द एक पृथक संकल्पना को स्पष्ट करता है। शब्दों के ये पर्याय-समूह संकल्पनीय अर्थों तथा शाब्दिक संबंधों के माध्यम से एक-दूसरे से जुड़े हुए होते हैं। एक वाक्य में कहें तो शब्दतंत्र अर्थों के माध्यम से सुव्यवस्थित किया गया मशीन के पढ़ने योग्य एक शाब्दिक संचय है। शब्दतंत्र की संरचना प्राकृतिक भाषा संसाधन के क्षेत्र में हो रहे शोधों के लिए भी उपयोगी साधन होती है। अर्थपूर्ण ढंग से जुड़े हुए शब्दों और



संकल्पनाओं के इस परिणामित तंत्र (Network) को विचरक (Browser) से देखा जा सकता है। अधिकतर शब्दतंत्र डाउनलोड करने के लिए सार्वजनिक रूप से निःशुल्क उपलब्ध हैं।

शब्दतंत्र शब्दों के मध्य विभिन्न शाब्दिक तथा अर्थ-संबंधी संबंधों को एक साथ लाने के लिए निर्मित तंत्र है। यह शाब्दिक जानकारियों को शब्दों के अर्थों के रूप में सुव्यवस्थित करता है तथा इसे मनोभाषाविज्ञान पर आधारित एक शब्दकोश कहा जा सकता है। हिंदी शब्दतंत्र शब्दों के मध्य विभिन्न प्रकार के संबंधों को दर्शाने का एक संगणकीय शाब्दिक तंत्र है। यह पारंपरिक शब्दकोशों से अलग है, जिसमें वैज्ञानिकता के आधार पर शब्दों के बारे में यथार्थ जानकारी उपलब्ध कराता है। इसकी संरचना अँग्रेजी शब्दतंत्र पर आधारित है पर यह हिंदी की विशिष्टता को समाहित किए हुए है।

हिंदी शब्दतंत्र में अर्थ की समानता के आधार पर पर्याय-समूह (Synset) का निर्माण एक स्पष्ट व्याख्या और वाक्य-प्रयोग के साथ किया जाता है ताकि शब्दों की अनेकार्थता के कारण अर्थ की स्पष्टता बाधित न हो। वास्तव में, ये पर्याय-समूह (Synset) ही हिंदी शब्दतंत्र के आधार हैं। फिलहाल अभी तक हिंदी शब्दतंत्र में संज्ञा, विशेषण, क्रिया एवं क्रिया-विशेषण वर्ग के शब्दों को ही स्थान दिया गया है।

इस शब्दतंत्र की संरचना में योगदान देने वाले प्रत्येक तत्त्वों से निम्न सीमाएँ हैं-

1. **पर्याय-समूह (Synset)** – इसमें किसी संकल्पना की यथार्थता को सूचित करनेवाले शब्दों को क्रम से बारम्बारता के आधार पर व्याख्या और उदाहरण के साथ दर्शाया जाता है। **जैसे-** गाय, गऊ, धेनु-सींगवाला एक शाकाहारी मादा चौपाया “गाय अपने बछड़े को दूध पिला रही है।”

इसमें तीन सिद्धांतों के आधार पर कार्य किया जाता है यथा-

- a) **संक्षिप्तता:-** कम से कम पर्याय-शब्दों के प्रयोग से संकल्पना की एकार्थता का स्पष्टीकरण। **जैसे-** घर, गृह।

- b) **व्यापकता :-** संकल्पना की एकार्थता को दर्शानेवाले सभी पर्याय-शब्दों को क्रम से उनकी बारम्बारता के आधार पर पर्यायवाची-समूह में प्रविष्टि। **जैसे-** घर, गृह, मकान, सदन, शाला, आलय, धाम, निकेतन, वास्तु, पण।
- c) **प्रतिस्थापनीयता:-** पर्याय-समूह में आए हुए अधिकतर शब्दों का वाक्य में बिना अर्थ परिवर्तन के प्रतिस्थापना संभव है- **जैसे-** घर, गृह, मकान, सदन, शाला, आलय, धाम, निकेतन, वास्तु, पण।

इनके आधार पर इस वाक्य को देखा जा सकता है यथा- मनुष्यों का छाया हुआ वह स्थान, जो दीवारों से घेरकर बनाया जाता है - "इस घर में पाँच कमरे हैं।"

2. सत्ता-मीमांसा (Ontology) – हिंदी शब्दतंत्र में शब्द-भेद के आधार पर शब्द की संकल्पना को सुस्पष्ट करने के लिए सत्ता-मीमांसा भी दी जाती है। **जैसे-** गाय, गऊ, धेनु

- पालतू पशु
- पशु
- सजीव
- संज्ञा

हिंदी शब्दतंत्र में संबंधों की रूपरेखा:- हिंदी शब्दतंत्र संकल्पना पर आधारित है एवं ये संकल्पनाएँ भाव की दृष्टि से प्रत्यक्ष या अप्रत्यक्ष रूप से एक-दूसरे से जुड़ी होती हैं। अस्तु इन संकल्पनाओं को एक-दूसरे से जोड़ने के लिए अधिवाची, अधोवाची, अंगवाची, अंगीवाची, विपर्यायवाची आदि संबंधसूचक अवधारणाओं का उपयोग किया जाता है। इनमें से कुछ शब्दों के मध्य के संबंधों को तो कुछ पर्याय-समूहों के मध्य के संबंधों को दर्शाती हैं।

1. अधिवाची (Hypernymy) एवं अधःवाची (Hyponymy):- पर्याय-समूहों के मध्य महत्वपूर्ण संबंध को दर्शानेवाली ये दोनों अवधारणाएँ एक दूसरे की पूरक हैं। यदि 'क' एक प्रकार का 'ख' है तो 'ख' 'क' का

अधिवाचक और 'क' 'ख' का अधःवाचक है और ये संबंध अधिवाची और अधःवाची कहलाते हैं। **जैसे-** गाय (क) ----- चौपाया (ख)

2. अंगवाची (Meronymy) एवं अंगी वाची (Holonymy):- ये संबंध सूचक अवधारणाएँ भी पर्याय-समूहों के मध्य होती हैं। यदि 'क' 'ख' का भाग है तो 'क' 'ख' का अंगवाचक और 'ख' 'क' का अंगीवाचक है और ये संबंध अंगवाची एवं अंगीवाची कहलाते हैं। **जैसे-** गाय (ख) ----- थन (क)

3. विपर्यायवाची (Antonymy):- यह संबंधसूचक अवधारणा पर्याय-समूहों के मध्य न होकर शब्दों के मध्य होती है। यह अवधारणा दो शब्दों के मध्य अवस्था, कार्य, समय, गुण आदि के आधार पर विपरीत अर्थ को दर्शाती है। **जैसे-** बेटा – बेटी, पुत्र – पुत्री

4. श्रेणीकरण (Gradation):- यह संबंधसूचक अवधारणा दो विपर्याय शब्द संकल्पनाओं के बीच की एक तीसरी शब्द संकल्पना को दर्शाती है। **जैसे-** सुबह – दुपहर – शाम

5. अपरिहार्यतावाची (Entailment):- यह संबंध दो क्रिया पर्याय-समूहों के मध्य होता है। यदि 'क' क्रिया 'ख' क्रिया में निहित है तो 'क' 'ख' का अपरिहार्यतावाचक है और यह संबंध अपरिहार्यतावाची कहलाता है। **जैसे-** 'खर्राटा लेना' क्रिया में 'सोना' क्रिया निहित है।

6. प्रकारवाची (Troponymy):- यह संबंध भी केवल दो क्रिया पर्याय-समूहों के मध्य होता है। इसमें एक क्रिया किसी दूसरी क्रिया के किसी विशेष ढंग को दर्शाती है। **जैसे-** 'मुस्कुराना' 'हँसना' का प्रकारवाचक है और यह संबंध प्रकारवाची कहलाता है।

7. प्रेरणार्थक क्रिया (Causative Verb):- इसमें मूल एवं प्रेरणार्थक क्रियाओं के मध्य संबंध दर्शाया जाता है। **जैसे-** चलना – चलाना, पढ़ना – पढ़ाना

शब्द-भेदों के मध्य संबंध:- हिंदी शब्दतंत्र में कुछ संबंध शब्द-भेदों के मध्य भी दिए गए हैं यथा-

1. संज्ञा-पद एवं क्रिया-पद के मध्य संबंध

A. आन्तर-योग्यता निर्देशी क्रिया (*Ability Link*):- इस संबंध के अन्तर्गत कोई क्रिया-पद किसी संज्ञा-पद के प्राकृतिक (मूल) गुण को दर्शाता है। **जैसे** –

- मछली, मीन, मत्स्य (संज्ञा-पद)
- तैरना, पैरना (क्रिया-पद)

B. बाह्य-योग्यता निर्देशी क्रिया (*Capability Link*):- इस संबंध के अन्तर्गत कोई क्रिया-पद किसी संज्ञा-पद के बाह्य या अर्जित योग्यता को दर्शाता है। **जैसे**–

- व्यक्ति, मानस (संज्ञा-पद)
- तैरना, पैरना (क्रिया-पद)

C. कर्म निर्देशी क्रिया (*Function Link*):- इस संबंध के अन्तर्गत कोई क्रिया-पद किसी संज्ञा-पद के कर्म को निर्दिष्ट करता है। **जैसे**–

- अध्यापक, शिक्षक (संज्ञा-पद)
- पढ़ाना, शिक्षा देना (क्रिया-पद)

2. संज्ञा-पद एवं विशेषण-पद के मध्य संबंध

A. गुणवाची (*Attribute*):- इस संबंध के अन्तर्गत कोई विशेषण-पद किसी विशेष संज्ञा-पद के गुणधर्म को सूचित करता है। **जैसे** –

- बाघ (संज्ञा-पद)
- मांसाहारी (विशेषण-पद)

B. अर्थ संकुचन-संज्ञा (*Modified Noun*) :- कुछ विशेषण-पद केवल कुछ विशेष संज्ञा-पद या संज्ञा-पदों को ही विशेषित करते हैं। ऐसे विशेषण-पद एवं संज्ञा-पद, अर्थ संकुचन-संज्ञा संबंध द्वारा एक दूसरे से जुड़े होते हैं।
जैसे-

- पंखदार, पाँखदार (विशेषण-पद)
- पक्षी, चिड़िया (संज्ञा-पद)

3. क्रिया-पद एवं क्रियाविशेषण-पद के मध्य संबंध :-

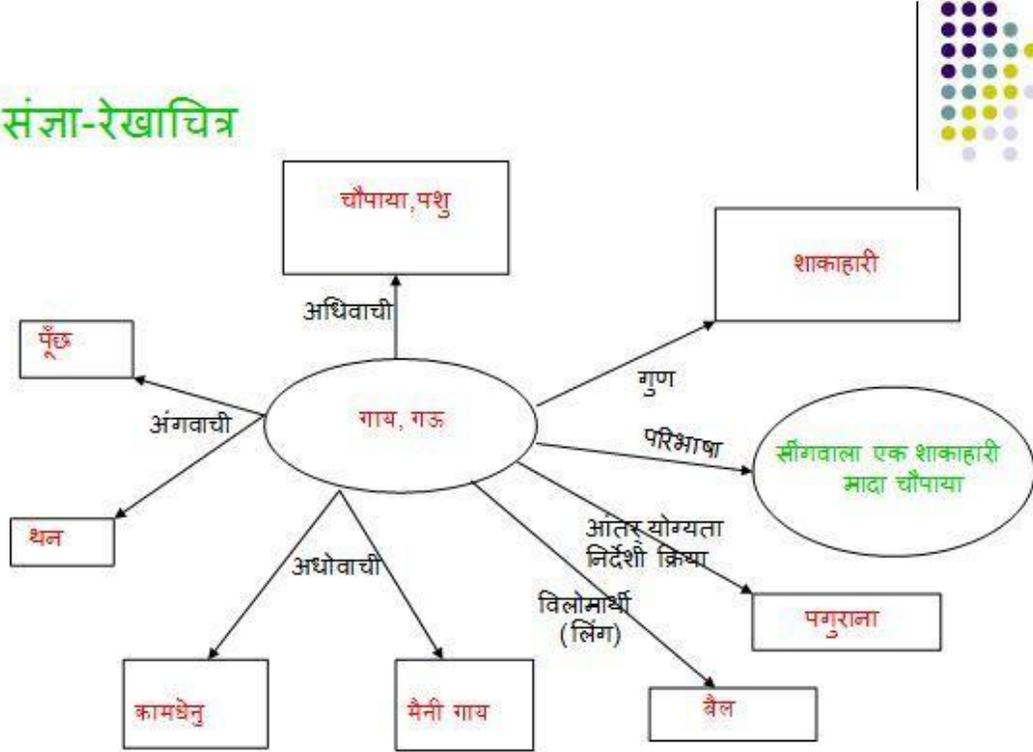
A. अर्थ संकुचन-क्रिया (*Modified Verb*):- कुछ क्रियाविशेषण-पद केवल कुछ विशेष क्रिया-पद या क्रिया-पदों को ही विशेषित करते हैं। ऐसे क्रियाविशेषण-पद एवं क्रिया-पद, अर्थ संकुचन-क्रिया संबंध द्वारा एक दूसरे से जुड़े होते हैं। **जैसे-**

- रिमझिम-रिमझिम, (क्रियाविशेषण-पद)
- बरसना, वर्षा होना (संज्ञा-पद)

B. से व्युत्पन्न (*Derived from*) :- इस संबंधबोधक अवधारणा द्वारा यह दर्शाया जाता है कि व्युत्पन्न शब्द की व्युत्पत्ति किस मूल शब्द से हुई है। **जैसे-**

- क्रमशः, क्रमानुसार, क्रमवार, सिलसिलेवार
- क्रम, सिलसिला

संज्ञा-रेखाचित्र



इस प्रकार, हम देख सकते हैं कि शब्द की अर्थ-स्पष्टता और विषय-वस्तु के सटीक वर्गीकरण के साथ-साथ शब्दों के मध्य के आपसी संबंधों को दर्शानेवाला हिंदी शब्दतंत्र भाषा और संगणक का एक चमत्कारी सुमेल है जो शब्दों का उचित विश्लेषण करके प्रयोगकर्ताओं के आगे अथाह शब्द-व्यंजनों को परोसता है।

हिंदी शब्दतंत्र का निर्माण-कार्य अभी भी 'भारतीय प्रौद्योगिकी संस्थान मुंबई' के 'संगणकविज्ञान एवं अभियांत्रिकी विभाग' में प्रो. पुष्पक भट्टाचार्या के मार्गदर्शन में अविराम चल रहा है। अभी 25/11/09 तक इसमें तैंतीस हजार छह सौ से ऊपर पर्याय-समूह (Synset) हैं, जो अस्सी हजार नौ सौ मूल शब्दों को समाहित किए हुए है। हिंदी शब्दतंत्र के लगभग तेरह हजार पाँच सौ प्रयाय-समूहों को अँग्रेजी शब्दतंत्र के इतने ही समतुल्य पर्याय-समूहों से जोड़ा भी जा चुका है, जो आनलाइन उपलब्ध हैं। अभी तक हिंदी शब्दतंत्र में लोगों द्वारा दो लाख पचहत्तर हजार पाँच सौ से ऊपर शब्दों की खोज की जा चुकी है तथा पाँच हजार एक सौ बासठ लोगों द्वारा इसको डाउनलोड किया जा चुका है। हिंदी शब्दतंत्र के आधार पर ही मराठी, संस्कृत, तमिल, तेलुगु, मलयालम,



कन्नड़, कोंकणी, नेपाली, मणिपुरी, बोड़ो, असमी और कश्मीरी में भी शब्दतंत्र का निर्माण किया जा रहा है। यह यूरोप (ELRA) एवं यूएसए (LDC) के सुप्रतिष्ठित प्राकृतिक भाषा संसाधन समूहों में शामिल है एवं साथ ही साथ विश्व के अन्य कई प्राकृतिक भाषा संसाधन समुदायों द्वारा उपयोग में लाया जा रहा है। यांत्रिक अनुवाद में तेजी लाने और सूचना-संचयन के लिए यह गूगल द्वारा भी अपनाया गया है। इसको 2008 के पी. के. पटवर्धन प्रौद्योगिकी विकास पुरस्कार से भी सम्मानित किया।

Citation: पाण्डेय, प्रभाकर & कश्यप, लक्ष्मी & भट्टाचार्य, पुष्पक (2010). हिंदी शब्दतंत्र की संरचना, HindiTech: A Blind Double Peer Reviewed Bilingual Web-Research Journal, 1 (1), 1-8. URL: <https://hinditech.in/hindi-shabdatantra-ki-sanrachana/>